9. *Probability Distribution*

In a probability distribution, the variables are distributed according to some definite probability function. Various types of probability distributions are as follows:

(I) Binomial Distribution

The Binomial distribution is also known as **"the outcome of Bernoulli process"** and is associated with the name of **"jacob Bernoulli."** A Bernoulli process is a random process in which :

(a) *The process is performed under the same conditions for a fixed and finite number of trials, say, n.*

(b) *Each trial is independent of other trials, i.e., the probability of an outcome for any particular trial is not influenced by the outcomes of the other trials.*

(c) *Each trial has two mutually exclusive possible outcomes, such as "success" or "failure", "good" or "defective", "yes" or "no", "hit" or "miss", and so on. The outcomes are usually called success and failure for convenience.*

(d) *The probability of success, P, remains constant from trial to trial (so is the probability of failure q, where, $q = 1 - p$).*

The general model for specifying the probability of obtaining exactly x successes in a given number of n, Bernoulli trials is given by

Focus Formula

$$f(x) = P[X = x] = {}^nC_x p^x q^{n-x} \text{ for } x = 0, 1, 2, ..., n.$$

where p = *the probability of a success on a single Bernoulli trial*

n = *the number of Bernoulli trials.*

x = *the number of successes in n trials.*

This formula for the probability distribution of the number of successes in series of Bernoulli trials is called the Binomial probability distribution. It gives the probability of obtaining exactly x successes and (n – x) failures in n Bernoulli trials.

The Binomial distribution satisfies the two essential properties of probability distribution, viz.,

(i) $f(x) \geq 0$; and

(ii) $\Sigma f(x) = 1$.

Mean and Variance of Binomial Distribution

The mean of binomial random variable X, is the theoretical expected number of successes in n trials. It is denoted by μ.

Focus Formula

$$\mu = E(X) = \sum_{x=0}^{n} x f(x) = \Sigma x\, {}^nC_x p^x q^{n-x}$$

$$\mu = np$$

Here n = *number of trials*

p = *probability of success in a single trial*

The variance of the binomial random variable X measures the variation of the binomial distribution. It is given as follows:

Focus Formula

$$\sigma^2 = E(X^2) - \mu^2 = \sum_{x=0}^{n} x^2 f(x) - \mu^2$$
$$= npq$$

Here $n = number\ of\ trials$

$p = probability\ of\ success\ in\ a\ single\ trial$

$q = probability\ of\ failure\ in\ a\ single\ trial$

Note: Binomial distribution may be negatively or positively skewed.

Example : The mean of a binomial distribution is 40 and standard deviation 6. Calculate n, p and q.

Solution : The mean of binomial distribution is given by np and standard deviation by $\sqrt{npq}$.

Since $\sqrt{npq}$. = 6; npq = 36 and np = 40

Therefore,

$$40q = 36 \text{ or } q = \frac{36}{40} = 0.9$$

$$p = 1 - q = 1 - 0.9 = 0.1$$

Given, $np = 40$ or $n = \frac{40}{p} = \frac{40}{0.1} = 400.$

Hence for the given question, n = 400, p = 0.1 and q = 0.9.

Example : Suppose that the half of the population of town are consumers of rice. One hundred investigators are appointed to find out its truth. Each investigator interviewed 10 individuals. How many investigators do you expect to report that three or less of the people interviewed are consumers of rice ?

Solution : Probability of a person being consumer of rice is p = 1/2, q = 1/2. Probability that three people or less are consumers of rice is given by

$$P[X \le 3] = P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3]$$

$$= q^{10} + {}^{10}C_1\, q^9\, p^1 + {}^{10}C_2\, q^8\, p^2 + {}^{10}C_3\, q^7\, p^3$$

$$= \left(\frac{1}{2}\right)^{10} + 10\left(\frac{1}{2}\right)^{10} + 45\left(\frac{1}{2}\right)^{10} + 120\left(\frac{1}{2}\right)^{10}$$

$$= \left(\frac{1}{2}\right)^{10} (1 + 10 + 45 + 120) = \frac{176}{1024}$$

Therefore, the number of investigators to report that three or less people are consumers of rice is given by

$$= \frac{176}{1024} \times 100 = 17.2 = 17 \text{ approx.}$$

(II) Poisson Distribution

The characteristics of the Poisson distribution are as follows :

1. The occurrence of the events is independent. That is, the occurrence of an event in an interval of space or time has no effect on the probability of a second occurrence of the event in the same, or any other interval.
2. Theoretically, an infinite number of occurrences of the event must be possible in the interval.
3. The probability of single occurrence of the event in a given interval is proportional to the length of the interval.
4. In any infinitesimal (extremely small) portion of interval, the probability of two or more occurrences of the event is negligible.

Poisson distribution differs from the binomial distribution in two important aspects :

(a) Rather than consisting of discrete trials, the distribution operates continuously over some given amount of time, distance, area, etc.

(b) Rather than producing a sequence of successes and failures, the distribution produces successes, which occur at random points in the specified time, distance, area. These successes are commonly referred to as 'occurrences'.

The Poisson distribution may be used to approximate binomial distribution when n is large and p is small and, therefore, is regarded as the limit of the binomial distribution.

The Poisson distribution is given by

$$\boxed{f(x) = P(X = x) = \frac{e^{-m} m^{x}}{x!};\ x = 0, 1, 2 \ldots}$$

where, m is called the parameter of the distribution and is the average number of occurrences of random event, x is the number of occurrences of the random event and e is the constant whose value is 2.7183.

The Poisson distribution satisfies the two essential properties, i.e., $f(x) \geq 0$ and $\Sigma f(x) = 1$.

Did You Know ?

The following are some of the examples which may be analysed with the use of this distribution:

(a) *the demand for a product,*
(b) *typographical errors occurring on the pages of a book,*
(c) *the occurrence of accident in a factory,*
(d) *the arrival pattern in a departmental store,*
(e) *the occurrence of flaws in a bolt in a factory, and*
(f) *the arrival of calls at a switch board.*

Mean and Variance of Poisson Distribution

The mean of the Poisson Distribution is given by

Focus Formula

$$\mu = E(X) = \Sigma x f(x) = \sum x \frac{e^{-m} m^x}{x!}$$

$$\mu = m = np$$

The variance of the Poisson Distribution is given by:

$$\sigma^2 = E(X^2) - [E(X)]^2 = E(X^2) - m^2$$

$$\sigma^2 = m = np$$

So in a Poisson Distribution mean and variance are equal.

Note : Poisson distribution is positively skewed.

Example : What probability model is appropriate to describe a situation where 100 misprints are distributed randomly throughout the 100 pages of a book ? For this model, what is the probability that a page observed at random will contain at least three misprints ?

Solution : Since there are 100 misprints in 100 pages, it implies that there is only one mistake on the average in a page. Therefore, the probability of being a misprint is very small as a page contains large number of words and n the number of words in 100 pages will be very large. So, in this case probability of being a misprint is small and n is very large, therefore, Poisson distribution is best suited here.

Average or expected number of misprints in one page is

$m = np = 100 \times 0.01 = 1$

$e^{-m} = e^{-1} = 0.3679$

Probability of at least three misprints in a page is

$= P[X \geq 3] = 1 - P(X < 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$

$$= 1 - \left[e^{-1} + e^{-1} + \frac{e^{-1}}{2!}\right]$$

$= 1 - e^{-1}\,[2.5] = 1 - 0.3679\ (2.5)$

$= 1 - 0.9198 = 0.0802.$

Fitting a Poisson Distribution: The process of fitting a Poisson distribution is simple. We have to obtain the values of m and calculate the probability of zero occurrences. The other probabilities can be easily calculated by the recurrence relation as follows :

$$f(x) = \frac{e^{-m} m^x}{x!},$$

and $$f(x+1) = \frac{e^{-m} m^{x+1}}{(x+1)!}$$

$$\frac{f(x+1)}{f(x)} = \frac{m}{x+1}$$

or $f(x + 1) = \frac{m}{x+1} f(x)$

when $x = 0$, $f(1) = m\, f(0)$

when $x = 1$, $f(2) = \frac{m}{2} f(1) = \frac{m^2}{2} f(0)$

when $x = 2$, $f(3) = \frac{m}{3} f(2) = \frac{m^3}{6} f(0)$

and so on.

This recurrence relation provides a very convenient method for fitting the Poisson distribution. The only probability we need to know is $f(0) = e^{m}$, where m is the only parameter of the Poisson distribution.

Multiplying by N (the total frequency) each probability of the Poisson distribution. We get the expected frequencies for respective probabilities.

(III) Normal Distribution

The Normal Distribution was discovered by De Moivre in 1733. It is an approximate to binomial distribution. Whether or not p is equal to q, the binomial distribution tends to the form of the continuous curve when n becomes large.

A random variable X is said to have a normal distribution with parameters μ (mean) and σ^2 (variance) if the density function is given by :

Focus Formula

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \qquad -\infty < X < \infty$$

where y = the computed height of an ordinate at a distance of X from the mean,

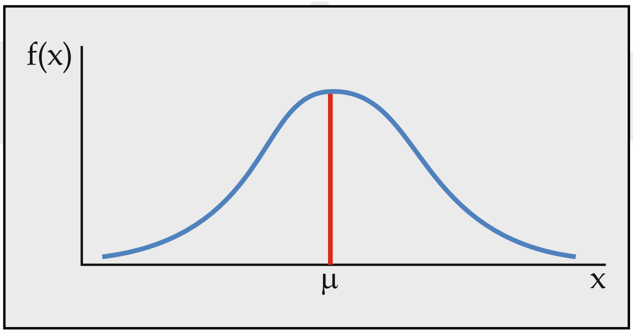
σ = *Standard deviation of the given normal distribution,*

π = *the constant* = 3.1416 ; $\sqrt{2x} = 2.5066$,

e = *the constant* = 2.7183 *(the base of the system of natural logarithm),*

μ = *Mean of the given normal distribution.*

If we draw the graph of normal distribution, the curve obtained will be known as normal curve and is given below.



The graph of y = f(x) is a famous 'bell shaped' curve. The top of the bell is directly above the mean μ. For large value of σ, the curve tends to flatten out and for small values of σ, it has a sharp peak.

The equation to a normal curve corresponding to a particular distribution is given by

$$\boxed{y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}}$$

The quantity $\frac{N}{\sigma\sqrt{2\pi}}$ in the above formula is equal to the maximum ordinate of the normal curve which will always occur at the mean of the distribution.

Relation between Binomial, Poisson and Normal Distribution

When n is large and the probability p of the occurrence of an event is close to zero so that np remains a finite constant, then the binomial distribution tends to Poisson distribution.

Normal distribution is a limiting form of Binomial distribution under the following conditions:

(a) n, the number of trials is very large, i.e. $n \to \infty$, and

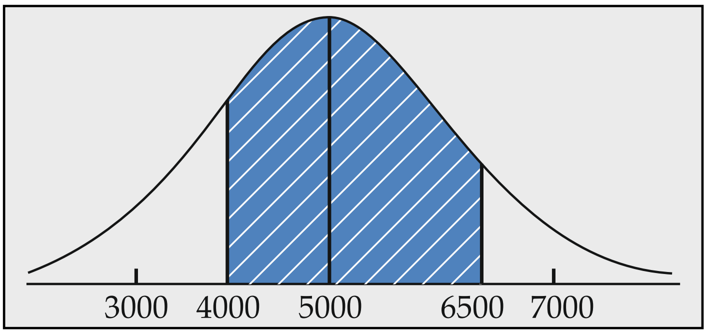
(b) neither p nor q is very small.

Properties of Normal Distribution

1. *The normal curve is symmetrical about the mean (skewness = 0).*
2. *The height of the normal curve is at its maximum at the mean.*
3. *There is one maximum point of the normal curve which occurs at the mean. The height of the curve declines as we go in either direction from the mean.*
4. *The normal curve is unimodal, i.e., it has only one mode.*
6. *The variable distributed according to the normal curve is a continuous one.*
7. *First and third quartiles are equidistant from the median.*
8. *Mean deviation about mean = 4/5σ or, more precisely, 0.7979 times of standard deviation.*
9. *Mean ± σ, mean ± 2σ, and mean ± 3σ covers 68.27%, 95.45%, and 99.73% area respectively.*

Example : How many workers have a salary between Rs. 4000 and Rs. 6500, if the arithmetic mean is Rs. 5000. Standard deviation is Rs. 1000 and number of worker is 15,000, if the salary of the worker is assumed to follow the normal law ?

Solution :



$$z_1 = \frac{4000 - 5000}{1000} = -1 \qquad \text{(left of the mean)}$$

$$z_2 = \frac{6500 - 5000}{1000} = 1.5 \qquad \text{(right of the mean)}$$

From the table, we find that 34.13% of workers fall between Rs. 4000 and Rs. 5000 and 43.32% fall between Rs. 5000 and Rs. 6500.

$\therefore$ 34.13 + 43.32 = 77.45% of workers have a salary between Rs. 4000 and Rs. 6500.

$\therefore$ Number of workers getting a salary Rs. 4000 and Rs. 6500 is given by

$0.7745 \times 15{,}000 = 11{,}618$

(IV) Exponential Distribution

A continuous random variable x is said to have an exponential distribution with parameter γ if its probability density function is given by

$$\begin{cases} P(x) = \lambda e^{-\lambda x} & \text{If } 0 \le x \le \infty, \lambda > 0 \\ \quad\;\; = 0, & \text{otherwise.} \end{cases}$$

The exponential variable takes values over an infinite range.

It can be proved that :

1. *The exponential density function decreases in the range 0 to ∞, the maximum ordinate of the curve occurring at x = 0 (the value being λ itself),*
2. *Larger the value of λ steeper is the decline in the ordinate, even for small values of x.*
3. *Smaller the value of λ, flatter does the curve become and lies closer to X-axis.*

The mean and variance of the exponential distribution are given by

$$\textit{Mean} = \frac{1}{\lambda}$$

$$\textit{Variance} = \frac{1}{\lambda^2}$$

Ques. *Which of the following relating to normal distribution are not correct ?*

(i) *Co-efficient of skewness is three.*

(ii) *It is mesokurtic.*

(iii) *Mean deviation for it is $\frac{2}{3}\sigma$.*

(iv) $\mu \pm 2\sigma$ covers 95.45% area.

(v) Mean, median and mode are equal.

(vi) The standard normal variate z has mean one and SD zero.

Codes : **(NTA UGC-NET June 2015 P-II)**

(A) (i), (iii) and (vi) (B) (iii), (iv) and (v)

(C) (i), (iii) and (v) (D) (i), (iii) and (iv)

Ans. (A)

- Co-efficient of skewness is three.
- Mean deviation for it is $\frac{2}{3}\sigma$.
- The standard normal variate z has mean one and SD zero.

Given above options are incorrect relating to normal distribution.